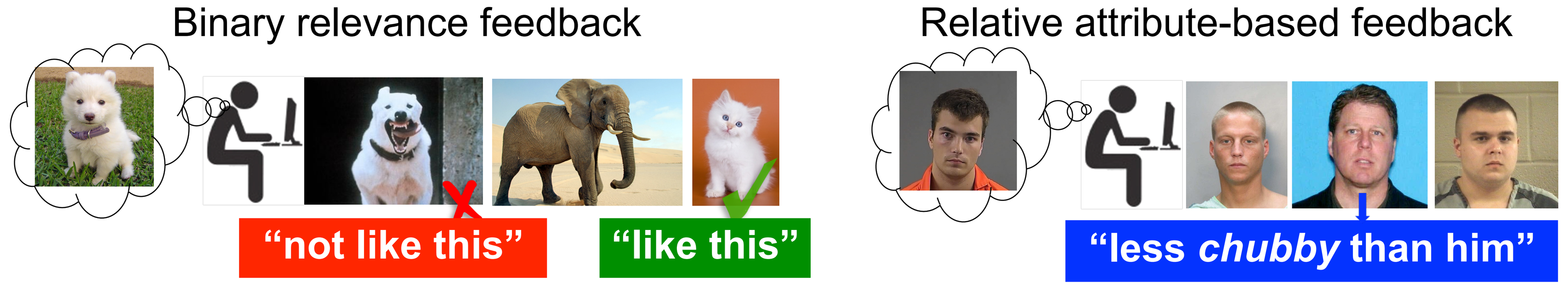
Implied Feedback: Learning Nuances of User Behavior in Image Search

Devi Parikh (Virginia Tech)   Kristen Grauman (UT Austin)

Intuition

In image search, a user's (perhaps subconscious) search strategy leads him to comment on certain images rather than others.



Feedback is a function of both the chosen image and the reference images the user sees but does *not choose to comment on.*

Key idea

- Whereas existing methods take user feedback at face value, we propose to learn the *implicit* information it conveys.
- We improve the efficiency of interactive image search by reading between the lines.

Approach

1. **Training:**
   a. Record interactions when people search for a target (known to us)
   b. Extract features revealing implicit selection biases
   c. Train relevance ranking function
2. **Testing:**
   a. Extract features from observed interaction
   b. Apply learned relevance ranking function
   c. Sort images based on likelihood of being the target image
   d. Iterate till user satisfied

Model: Learning a relevance ranking function

We learn a relevance ranking function S that accounts for implied feedback

True target   Distractors

$$S(\boldsymbol{t}_l) > S(\boldsymbol{x}_i)$$

$$\boldsymbol{w}^T\phi(\boldsymbol{t}_l, \Omega_l) > \boldsymbol{w}^T\phi(\boldsymbol{x}_i, \Omega_l)$$

Parameters to be learnt   Features characterizing interaction l

Max-margin learning to rank formulation

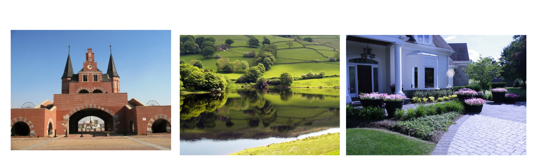
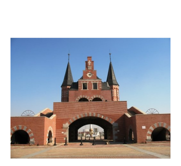
$$\min_{\boldsymbol{w},\xi_{il}} \frac{1}{2}||\boldsymbol{w}||_2^2 + C\sum \xi_{il}^2$$

$$\text{s.t. } \boldsymbol{w}^T\phi(\boldsymbol{t}_l, \Omega_l) \geq \boldsymbol{w}^T\phi(\boldsymbol{x}_i, \Omega_l) + 1 - \xi_{il}$$

$$\forall \boldsymbol{x}_i \neq \boldsymbol{t}_l, \forall l, \quad \xi_{il} \geq 0.$$

[Joachims 2002]

We represent an interaction with a 4-tuple

$\Omega_l = \langle$  ,  , natural , +1 $\rangle$

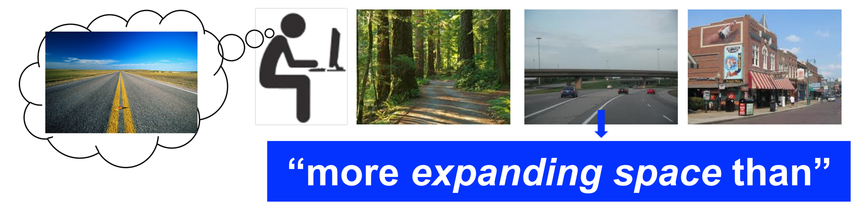
available reference images, chosen reference, chosen statement (like/not-like or attribute), polarity of statement (constant for binary relevance)

Features revealing implicit search strategies

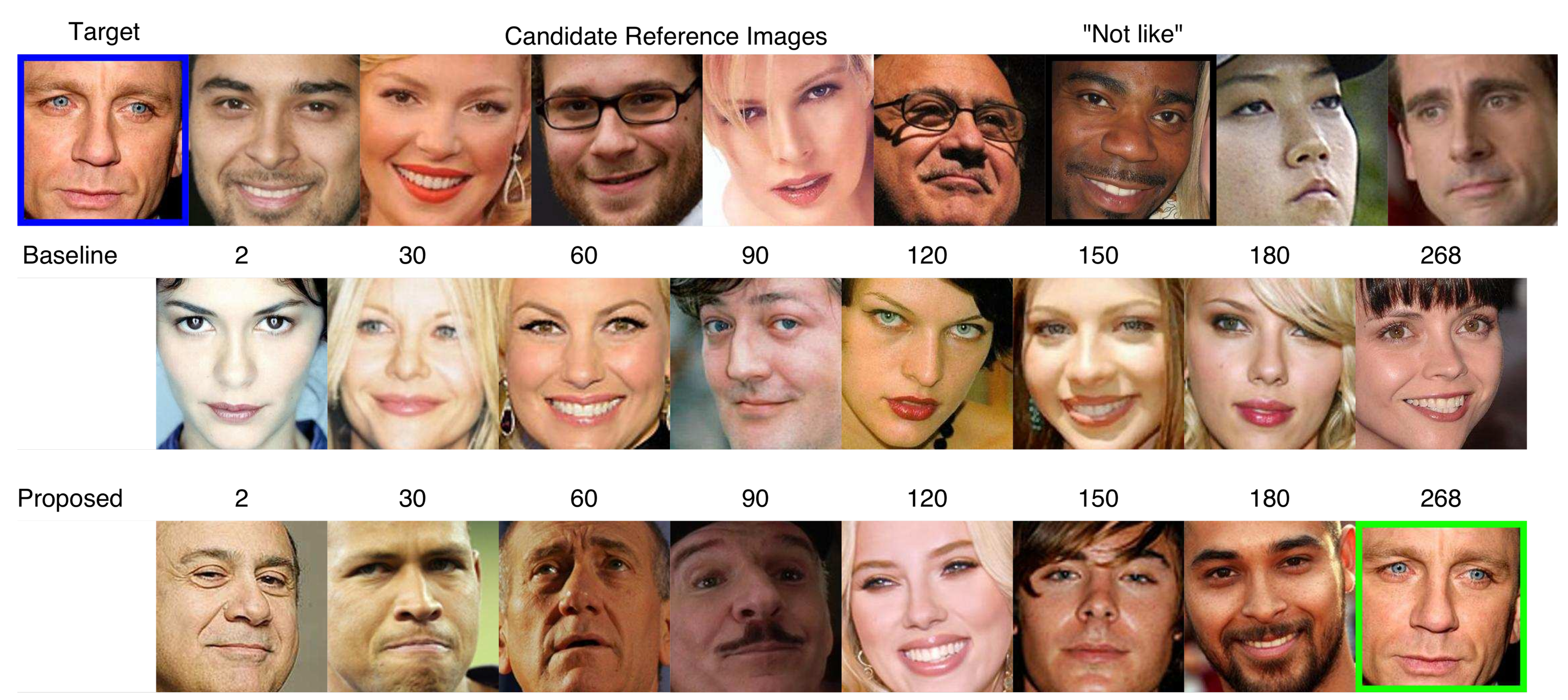
We introduce an array of features $\phi(\boldsymbol{t}_l, \Omega_l)$ to capture the implicit user reactions, based on relationships between the selected and non-selected reference images.

Binary relevance feedback:
- Distance of selected reference image from target image
- Relative to distance of other reference images from target
- Relative to visual diversity of reference images
- Variations (total 5 features)

Relative attribute-based feedback:
- Whether target image satisfies user-specified constraint or not
- How comfortably the constraint is satisfied
- "Tightness" of specified constraint
- Similarity of selected reference to target w.r.t chosen attribute
- Relative to similarity along other attributes
- Variations (total 31 features)



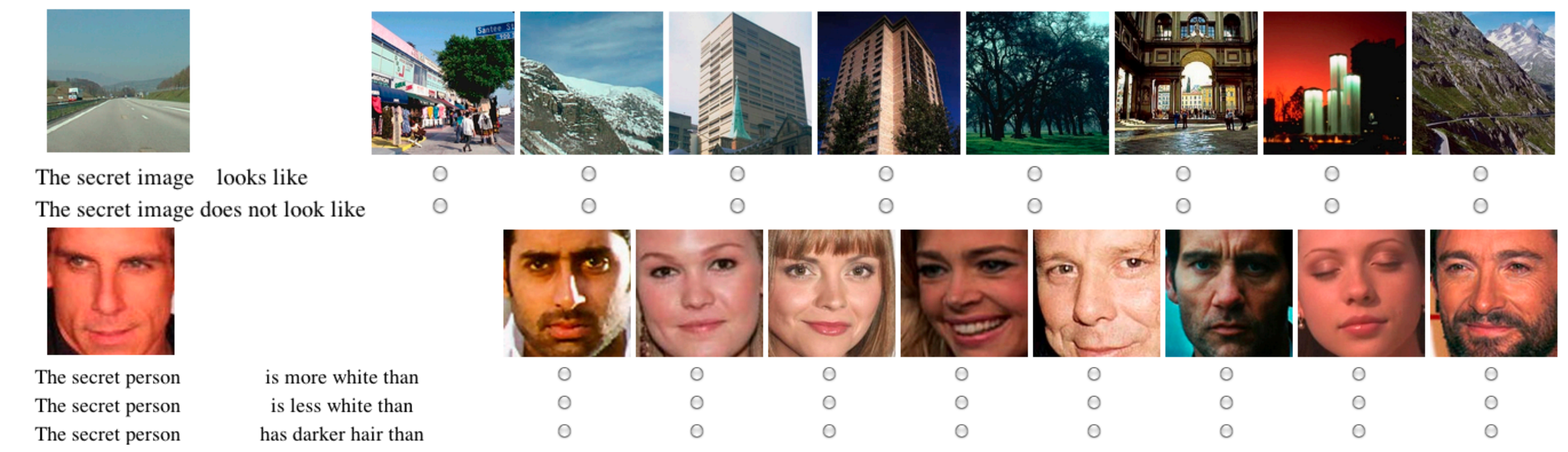
Qualitative results



We infer what's "behind" the user's feedback, learning from both what he says and doesn't say. As a result, we more rapidly converge on his target content.

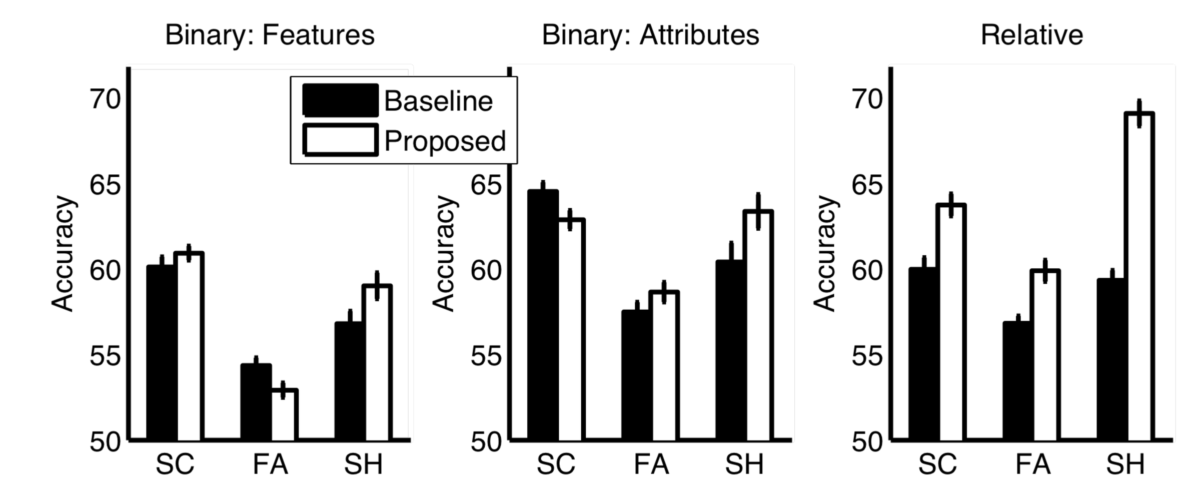
Data collection

Scenes (2688 images, 3 attributes), Faces (900 images, 10 attributes), Shoes (1000 images, 10 attributes).
Amazon Mechanical Turk, 1200 interactions, ~60 subjects

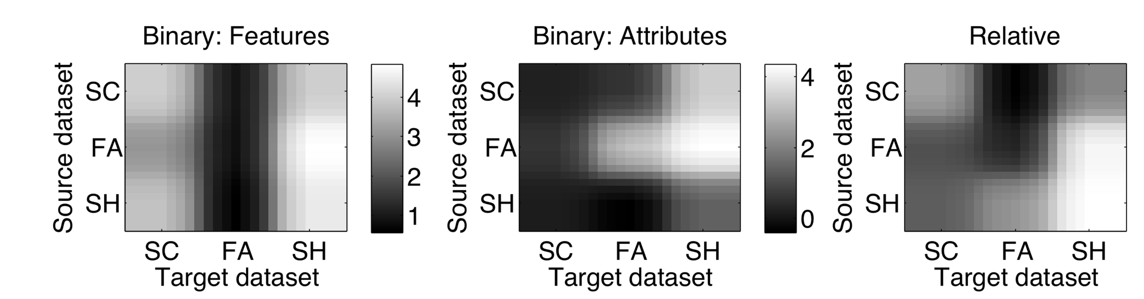


Results

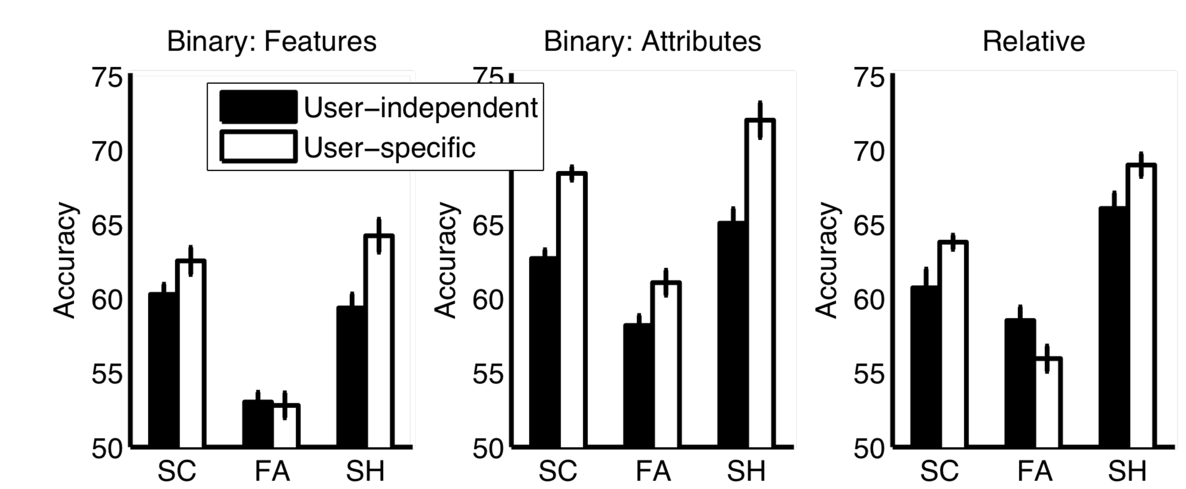
Comparison to traditional feedback processing



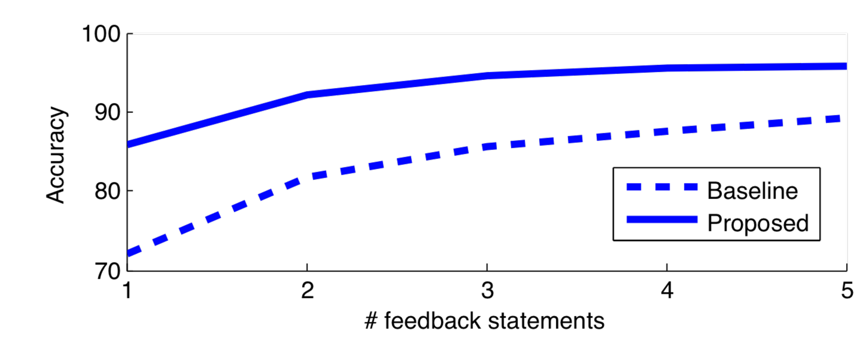
Do the implied cues generalize across domains?



Can we learn user-specific behavior?



Multiple feedback statements



Conclusion

✓ Implicit cues are embedded in existing forms of feedback
✓ We expose and leverage them for interactive image search
✓ Better accuracy, yet no additional overhead for user
✓ Results on multiple datasets with online image search users show clear impact